

REINAS: the Real-Time Environmental Information Network and Analysis System

Darrell D. E. Long, Patrick E. Mantey, Craig M. Wittenbrink,
Theodore R. Haining, Bruce R. Montague[†]
University of California, Santa Cruz

Abstract

The Real-Time Environmental Information Network and Analysis System (REINAS) is a distributed system supporting the conduct of regional environmental science research at the desk top. Continuous real-time data acquired from dispersed sensors is stored in a logically integrated but physically distributed data base. An integrated problem-solving environment is under development which supports visualization and modeling. REINAS is intended to provide insight into historical, current, and predicted oceanographic and meteorological conditions. REINAS permits both collaborative and single-user scientific work in a distributed environment.

Keywords: distributed data base, heterogeneous data base, real-time data base, temporal data base, scientific data base, geographical information system, environmental monitoring.

1 Introduction

The Real-Time Environmental Information Network and Analysis System (REINAS) supports regional-scale environmental science, monitoring, and forecasting on a distributed system connected by the Internet. It is being developed by the University of California at Santa Cruz (UCSC), the Naval Postgraduate School (NPS), and the Monterey Bay Aquarium Research Institute (MBARI).

REINAS is designed to provide on the desk top:

- A set of tools to configure and collect data from instruments in the field in real time,
- An integrated problem solving and visualization system supporting individual and collab-

orative research using both historical and modeled data, and

- A logically consistent distributed data base that stores data in the data base independent of file format and which maintains metadata describing where and how data was obtained (the data base tracks data pedigree).

Professionals in the environmental sciences will have the ability to observe, monitor, and analyze regional oceanographic and meteorological phenomena from their desk top.

REINAS serves different user groups: *Operational forecasters* monitor current conditions, view standard data products, synthesize new data views, and issue forecasts and warnings. *Modelers* analyze new model products and compare them with other models and with past and present conditions. *Experimental scientists* collaborate with other scientists on-line, observe individual data fields as they are collected, and may modify data collection methods for an ongoing experiment. Finally, *instrument engineers* add new equipment to the system, access metadata describing individual devices and methods of calibration, study maintenance records, and profile sensor quality.

Interactive real-time measurement, real-time monitoring, and retrospective data management using REINAS provides forecasters and experimental scientists with a framework to plan their experiments, expeditions, and monitoring activities. The special focus on real-time observation and analysis allows the oceanographic and meteorological communities to identify phenomena as they occur and to react to emerging phenomena and trends by changing the nature and frequency of data collection at sites of interest.

The integrated problem solving and visualization environment provides scientists and forecasters with pictures of environmental features, trends,

[†]Supported in part by the Office of Naval Research under Grant N00014-92-J-1807.

and relationships to study dynamic environmental behavior in a geographic context. Scientists will be able to fuse data collected in the field with data generated by numerical models and simulations.

System configuration tools simplify instrument engineering tasks, for instance, supporting easy addition of new instruments to the system. Instruments can be connected to REINAS by both radio and terrestrial links.

The data base supports access to data collected by different people and institutions at different times in a logically consistent manner. REINAS can collect data continuously in real-time. The data base separates users from many of the mechanical tasks of data management, as it is designed around an architecture integrating data from multiple instrument technologies, classes (numeric, text, and video), and representations. Both scientific data and metadata are stored using a stable schema which maintains data pedigree. An interactive electronic log book allows experimental scientists to keep notes easily accessible.

2 The Scientific Challenge

In real-time environmental science, data collection, data delivery, and data transformation activities feed the iterative processes of data analysis, modeling, and visualization. This information life cycle (see Figure 1) begins when instruments produce basic measurements. Such raw measurements are often inaccurate or in units meaningless in a broad context (a thermistor voltage representing temperature, for example). Calibration algorithms must then be applied to measurements to obtain usable observations. Observations are often sparse and irregular and must be assimilated into a consistent representation of the phenomena being observed. Models are often used iteratively to support data assimilation and in generating forecasts. Finally, visualization is used to present data. Observations and data are often useless to other researchers since data set pedigree, calibration algorithm information, and specific sampling method information are unavailable.

Our sources produce measurements in many formats including land based weather stations, ships, buoys, aircraft, radars, and satellites. Data from any of these data sources must be exchanged between systems using data exchange standards adopted by national and international weather centers. While observations from many sources are provided by

operational networks, these data is often not real-time and delivery often involves manual operations. Thus, researchers miss opportunities to coordinate data collection as phenomena occur.

3 REINAS System Architecture

REINAS integrates group activity over the data life cycle by supporting a common information model tracking individual data element and system resource lineage. The architecture provides the following benefits:

- Access to real-time and retrospective data,
- The performance required to support the most frequently requested products and services,
- Access to resources and devices through a common data model supporting rapid system configuration,
- Dynamic control of devices for real-time interactive scientific investigation,
- Fault tolerant data collection avoiding data loss due to communication link failures, and
- Security features restricting user access and control with respect to data and equipment.

3.1 System Organization

As shown in Figure 2, REINAS consists of three subsystems. The *instrumentation subsystem* collects data in a standard portable format. The *data base subsystem* stores data and provides a framework for data manipulation. The *user subsystem* provides users access to the instrument and data base subsystems and supports visualization and modeling applications. Each subsystem has its own logical network defined by ownership and operational characteristics. Functionality is distributed throughout these networks, with computation occurring at the most appropriate location.

The instrumentation subsystem nodes contain sensors attached to PCs running BSD/386¹ and connected to the Internet. Each PC converts data to a common format, writes the data to a local log, and transmits the data to data base nodes. Configuration and control functions for specific instruments are executed on the PC.

¹A Unix-like operating system distributed by Berkeley Software Design Inc., Colorado Springs, CO.

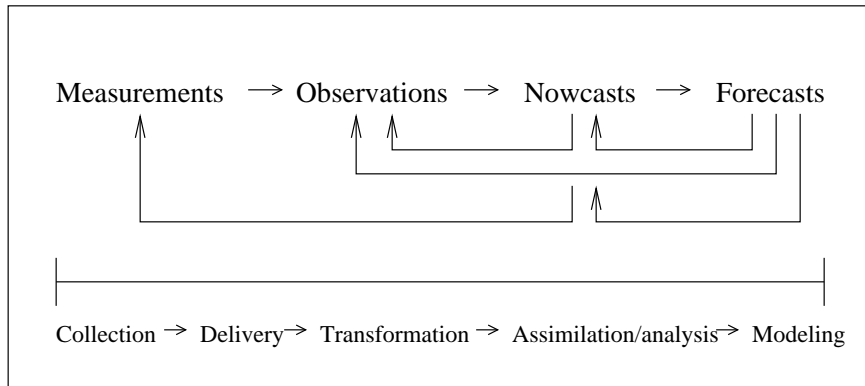


Figure 1: Environmental information life cycle – products and processes.

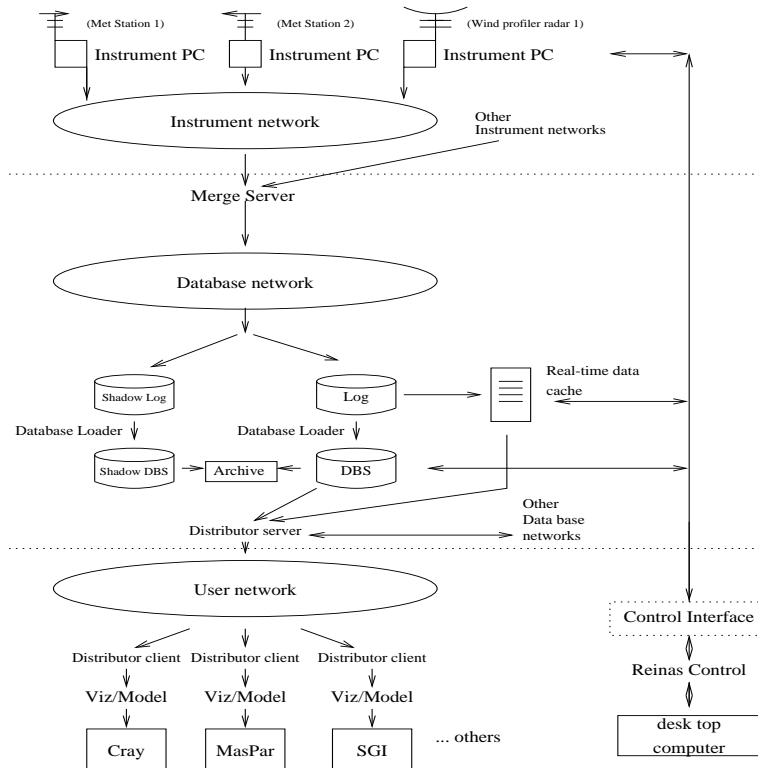


Figure 2: REINAS Logical Architecture

Data transmitted to a data base node is received by a *merge server* process and written to a data base load log. A *data base loader* process then reads the log and writes into the data base. Data is also read from the log into a memory resident real-time cache which contains the most current data values. User applications access data from both the historical data base and the real-time cache using a REINAS run-time library which communicates with a data base node *distributor* process.

The REINAS application run-time library implements a REINAS API (Application Programming Interface). The API provides a standard programming model for accessing and controlling both data base and instrument subsystems and for retrieving data. The run-time library also contains visualization support routines.

REINAS stores individual data elements in relational data base tables along with sufficient information to locate related data. Many data feeds, including real-time instrumentation, produce data in simple file formats. The diversity of such file formats create a considerable operational problem for scientific software. An important goal of the REINAS data base schema design effort has been to subsume these file formats.

The data base is fundamentally organized around a set of data time-series. The data elements of each unique scientific data type are stored in a table specific to that type. Data elements are primarily indexed by time regardless of the location from which the data was obtained. Each data element is tagged with its location. This design supports the rapid playback and display of data in the same temporal order in which it was obtained in real-time.

The data base *load path* consists of all the processing that must be performed between the raw data feed and the data base. New devices are added to REINAS by writing a *device manager*. Device managers implement a specific load path and are similar to device drivers in that they follow a standard documented framework and convert raw I/O into data recognizable to the system. A load path programmer must write five functions in a specified manner to implement a new device manager. The typical device manager so far requires between two thousand to four thousand lines of code.

REINAS treats data base systems as SQL engines. Any SQL data base can be used by a REINAS system by writing a REINAS *data base driver* which implements a small set of standard data base functions.

3.2 Advantages of the Architecture

Given this architecture, resources can be duplicated throughout the system and subsystems can communicate with multiple instances of other subsystems in higher and lower layers. For instance, instruments can provide input to multiple data base subsystems, and a data base subsystem can receive input from more than one instrument subsystem.

The design of each subsystem is optimized toward a specific task. The data base subsystem is optimized for data manipulation and storage through the use of the Andrew File System (AFS), large storage devices, and large data base servers. Computationally expensive applications can run on specialized compute nodes without impacting other nodes in the system.

Separation of concerns between layers makes the system extensible through the development of regular interfaces between machines on different layers. Since the layers communicate through the Internet, any machine on the Internet can be added through a protocol conforming to the appropriate interface.

REINAS contributes to scientific productivity by eliminating the need for each researcher to invent their own file handling routines. REINAS effectively does any data parsing that is required during the first steps of the data load process. Thus, researchers that implement a scientific application using the REINAS API effectively have direct access to the data with which they work in binary format. Unlike file specific applications, REINAS applications should continue to be of value regardless of the specific data feed or format.

The time-series data base design supports rapid multimedia style playback of recorded events. As data values are retrieved in temporal order they can be overlaid on spatial displays.

The REINAS device manager implements a real-time load path, as opposed to the manual load path often used in conventional scientific data bases. We hope to build up a library of device managers covering a number of standard device and data feed types.

Conventional scientific air/ocean instruments are often supported by small embedded data acquisition computers called *data loggers*. Data loggers often store information for long periods while the data is acquired. With REINAS, any Internet node can be used as a data logger which provides interactive real-time access and instrument control.

The REINAS data base drivers provide a common SQL interface to several different data base products. For example, the same REINAS application program

can thus access data in either an Oracle or Illustra data base engine. This is a non-trivial capability, as numerous small differences exist between SQL implementations, for instance, in areas such as date and time formats.

Many of the advantages of the REINAS system result from the use of a common application API on top of different data bases, networks, and operating systems. The term *middleware* has recently been applied to software of this nature, as it occupies the 'middle' between applications and specific operating systems, data bases, and networks.

3.3 Status

REINAS currently obtains real-time measurements from meteorological stations, wind profiler radars, CODARS (ocean surface current radars), and acoustic Doppler current profilers ([8]).

The REINAS data base currently exists in three alpha-level implementations for the Ingres, Illustra, and Oracle data base management systems (DBMS) [8]. All three data bases use the same base schema. Current engineering work centers on developing administration, browsing, and DBA tools using Tcl/Tk and Motif [10, 9], automating the device manager development process, developing device managers for additional devices, and developing a consistent API for the visualization environment. A surprisingly diverse number of time representations are used by scientific instrumentation. We hope to develop a single library of time manipulation functions that satisfy our needs. Model data remains to be integrated into the data base. Long range plans include expanding the data base subsystem into a multidata base to provide REINAS users with data managed by different organizations using different DBMS.

A prototype visualization environment containing approximately 40 thousand lines of C code has been developed on SGI platforms using IRIS GL [11, 12, 8]. A collaborative visualization extension allows scientists to build visualizations in a shared workspace [13].

The REINAS system software has been implemented in portable ANSI C. The system software is currently some 100 thousand lines of code in C and SQL and runs on Sun, SGI, DEC Alpha, IBM AIX, HP and BSD/386 (PC) workstations.

4 Related Work

The REINAS system is a large effort incorporating many research disciplines. Distributed computer technology is currently being used by projects such as SEQUOIA 2000 to support large environmental data bases [16, 1]. Systems such as BADGER are being built which focus on coastal or regional air/ocean science and apply geographical information systems (GIS) and visualization techniques [7, 2, 6]. REINAS differs from SEQUOIA 2000 and BADGER in its emphasis on real-time desk-top experimentation and its regional focus. REINAS expands on the approach of other systems by providing a common integrated data model to enhance the value of the scientific data collected.

Developing data base schemata tracking the meta-data associated with a scientific enterprise or scientific discipline is an ongoing activity [3]. The REINAS model supports the oceanographic and meteorological communities and has broad environmental application.

REINAS applies and extends many traditional computer science techniques. Distributed system techniques are applied in the areas of security, naming, and client-server application design [4, 14]. Although designed to be implemented using any conventional relational data base supporting SQL, REINAS uses an object-oriented data base design specific to the problem domain [5, 15].

5 Conclusions

The REINAS system has been designed for regional real-time environmental monitoring and analysis. REINAS is a modern system for conducting interactive real-time coastal air/ocean science on the Internet. The data base design of REINAS is independent of specific data base technology and is designed to support operational scientific needs throughout the entire scientific data life-cycle. REINAS supports extensibility through a clearly defined API, device management utilities, and data access tools. Current user applications support real-time and retrospective visualization of environmental data.

Acknowledgements

We would like to thank Bruce Gritton (MBARI), Professor Wendell Nuss and Dr. Paul Hirschberg

(NPS) for their valued collaboration, as well as our colleagues Professors J.J. Garcia-Luna, Alex Pang, and Harwood Kolsky. We are indebted to everyone involved in the REINAS project, notably Dr. Dan Fernandez and Andrew Muir.

References

- [1] Gustavo Alonso and Amr El Abbadi. Cooperative modeling in applied geographic research. Technical report TRCS 94-02. University of California - Santa Barbara, March 1994.
- [2] M. Botts, J. D. Dykstra, L. S. Elson, and S. J. Goodman. Challenges and opportunities in visualization for NASA's EOS Mission to Planet Earth. *Visualization '94* (Washington, DC, USA, 17-21 October 1994). Published as R. D. Bergeron and A. E. Kaufman, editors, *Proceedings. Visualization '94*, pages 392-5. IEEE Computer Society Press, 1994.
- [3] J. B. Cushing, D. Maier, M. Rao, and D. Abel. Computational proxies: modeling scientific applications in object databases. *Seventh International Working Conference on Scientific and Statistical Database Management* (Charlottesville, VA, USA, 28-30 September 1994). Published as J.C. French and H. Hinterberger, editors, *Proceedings. Seventh International Working Conference on Scientific and Statistical Database Management*, pages 196-206. IEEE Computer Society Press, 1994.
- [4] John H. Howard, Michael L. Kazar, Sherri G. Menees, David A. Nichols, M. Satyanarayanan, Robert N. Sidebotham, and Michael J. West. Scale and performance in a distributed file system. *ACM Transactions on Computer Systems*, 6(1):51-81, February 1988.
- [5] A. R. Hurson, Simin H. Pakzad, and Jia-bing Cheng. Object-oriented database management systems: Evolution and performance issues. *IEEE Computer*, pages 48-60, February 1993.
- [6] Smart Valley Inc. and Lockheed Research Laboratory. BADGER: Bay Area Digital GeoResource Executive Summary, 1994.
- [7] Ben Kobler and John Berbert. NASA earth observing system data information system (EOS-DIS). *Digest of papers, 11th IEEE Symposium on Mass Storage Systems* (Monterey, CA), pages 18-19, 7-10 October 1991.
- [8] P. E. Mantey, D. D. E. Long, J. J. Garcia-Luna, A. T. Pang, H. G. Kolsky, B. R. Gritton, and W. A. Nuss. REINAS: Real-Time Environmental Information Network and Analysis System Phase IV - Experimentation. Technical report UCSC-CRL-94-43. University of California - Santa Cruz, 1994.
- [9] OSF/Motif Series, Open Software Foundation, editor. Prentice Hall, 1992. 5 volume set.
- [10] John K. Ousterhout. An X11 Toolkit Based on the Tcl Language. *USENIX Conference Proceedings* (Dallas, TX), pages 105-15. USENIX, Winter 1991.
- [11] A. Pang and K. Smith. Spray rendering: Visualization with smart particles. *Visualization'93 Proceedings*, pages 283-90, 1993.
- [12] Alex Pang, Jeff Furman, and Wendell Nuss. Data quality issues in visualization. Pages 12-23. SPIE, February 1994.
- [13] Alex Pang, Craig M. Wittenbrink, and Tom Goodman. CSpray: A collaborative scientific visualization application. *Proceedings SPIE IS & T's Conference Proceedings on Electronic Imaging: Multimedia Computing and Networking*, to appear 1995.
- [14] Marc Shapiro. Structure and encapsulation in distributed systems: the proxy principle. *Proceedings of 6th International Conference on Distributed Computing Systems* (Cambridge, Mass), pages 198-204. IEEE Computer Society Press, Catalog number 86CH22293-9, May 1986.
- [15] M. R. Stonebraker and L. A. Rowe. The POSTGRES data model. *Thirteenth International Conference on Very Large Data Bases: 1987 13th VLDB* (Brighton, UK, September 1-4, 1987). Published as P. M. Stocker, W. Kent, and P. Hammersley, editors, *Proceedings of the Thirteenth International Conference on Very Large Data Bases: 1987 13th VLDB*, pages 83-96. Morgan Kaufmann, 1987.
- [16] Michael Stonebraker and Jeff Dozier. Large capacity object servers to support global change research. Technical report 91-1. SEQUOIA 2000, July 1991.